NASA-CR-203240

/NTERIM
IN-47-CR
OCIT
027237

SEMI-ANNUAL PROGRESS REPORT

Period: July-December 1996

"The Effects of Cloud Inhomogeneities Upon Radiative Fluxes, and the Supply of a Cloud Truth Validation Dataset"

Submitted by

Institute of Atmospheric Sciences
South Dakota School of Mines and Technology
501 E. St. Joseph Street
Rapid City, SD 57701-3995

to

National Aeronautics and Space Administration
Goddard Space Flight Center
Contract NAS5-31718
Greenbelt, MD 20771

1. Work Accomplished During the Report Period

This work is directed towards the development of algorithms for the ASTER science/instrument teams. Special emphasis is being placed on a wide variety of cloud optical property retrievals, and especially retrievals of cloud and surface properties in the polar regions.

2. Research Activities

2.1 Cloud Algorithms

2.1.1 ASTER Polar Cloud Mask

During this reporting period we found, from the testing performed during the first half of the year, that the hierarchical neural network (HNN), overall, performed better than the paired histogram method (PHM). This conclusion is based on results compiled from applying the classifiers to our set of labeled samples and generating confusion matrices from the results. Both classifiers performed equally well in distinguishing between cloudy and clear areas (approximately 95 percent accuracy); however, the HNN was more accurate in distinguishing among the four cloud classes (e.g., thin cloud over snow/ice, water, or land and thick cloud) and among the six clear classes (e.g., water, slush/wet ice, snow/ice, shadow on snow/ice, land, and shadow on land). The HNN was also found to be significantly more robust when partitioning the samples into mutually exclusive training and testing sets. The HNN classification performance decreased a few percent overall while the PHM classifier decreased by about 10 percent. The results from the testing of the HNN were assembled into a journal paper entitled "Landsat TM Polar Cloud Mask. Part I: Application of a Hierarchical Neural Network" and submitted to *IEEE Transactions on Geoscience and Remote Sensing* during the month of August.

Due to the improved performance observed when using the neural network approach, we subsequently integrated the first classification stage of the algorithm (now called the preclassifier), with a set of specially trained back propagation neural networks. The preclassifier is now designed only to reduce the classification ambiguity thereby narrowing the classification choices from 10 to something less than 10 classes. For example, the preclassifier might indicate that a pixel is a member of one of three cloud classes but not any of the other seven classes. The feature vector is then passed on to a neural network that is specially trained to distinguish between only the three cloud classes. In effect, what was formerly called the decision network (in previous reports) of the HNN has been supplanted by the preclassifier. The reason the preclassifier is retained in favor of the decision network is that the preclassifier is adaptive; that is, the thresholds used in the preclassifier tests are derived from the scene statistics and are not static across all the scenes. In the HNN, the same decision network is applied to all scenes in the same way. This new adaptive/hybrid neural network classifier has been applied to approximately 20 scenes and the classification results appear to be noticeably better than results obtained from past versions of the classifier. We are currently in the process of applying the classifier to our sample set for the purpose of generating confusion matrices. A journal paper describing this technique currently is being prepared and we plan to submit it early next year.

As described in several previous reports, in a parallel effort, a fuzzy logic classifier is being developed and tested. When applying the most recent version of this classifier to our sample set, the confusion matrices indicate that it performs nearly as well as the HNN. It also appears to provide good performance when applying it to full scenes. Due to its good performance on our current Landsat TM polar data set, another journal paper entitled "Landsat TM Polar Cloud Mask. Part II: Application of a Fuzzy Logic Classifier", was prepared describing the technique and the results obtained. The paper was submitted to *IEEE Transactions on Geoscience and Remote Sensing* during the month of November. The paper describes another significant aspect of this development, aside from its use as a classifier, and that is as a feature reduction technique. Like the neural network classifier, it is impractical to utilize 140 features in the classifier in an operational environment; therefore, the length of the feature vector must be reduced as much as possible and certainly to less than 40. In this effort, features were selected through a forward and backward elimination process as follows. In the forward elimination process, one feature at a time is added to the feature set and then the fuzzy logic classifier is applied to the sample set. If the confusion matrix derived from the results indicates that the addition of the feature improved performance, the feature is retained in the feature vector. If the performance decreases or does not change, then the feature vector is not added to the feature vector. Similarly, in backward elimination, initially all features are retained in the feature vector. Features are removed one at a time from the feature vector and the confusion matrix is examined. If the performance increases or does not change, the feature is not returned to the feature vector. If the accuracy decreases, the feature is retained in the feature vector. Using this process, it was found that only 20-25 features are required to achieve the best performance from the classifier. We are now testing the neural network classifier using the feature sets derived from this process and have observed no decrease in performance from this reduced feature set (the feature set previously numbered 39). We will now be using this feature selection process in the algorithm development, which will replace the older method based on distributional overlap and divergence.

In the preclassifier some simple adaptive and simple fixed thresholds are used in a class ambiguity reduction process. For example, as a result of a few tests the possible classification choices for a pixel feature vector is reduced from 10 classes to 2-4 classes. The neural network is then greatly simplified as the number of inputs and outputs is reduced. Many of the current simple thresholds applied in this stage were found through analysis of scatter plots, feature histograms, and physical phenomenology. Sometime after launch, when registered and calibrated ASTER data becomes available, it will be necessary to adapt the algorithm from the developmental Landsat TM dataset to that of ASTER. The following paragraphs discuss a 6 month Post Launch Operational Plan to accomplish this. To facilitate the adaptation of the preclassifier to the ASTER data, we are developing some automated techniques for deriving a set of simple thresholds that will perform the class ambiguity reduction. Classes are being examined in various groupings (e.g., clear vs. cloud) to find a set of thresholds that provide for the most reliable and robust reduction for the largest fraction of pixels. Once this process is automated the benefits should be three-fold. First, the performance should improve as more reliable class ambiguity reduction is provided. Second, the classification speed should increase as the tests that provide for the most reduction will be used. And thirdly, the tests will be derived in an automatic and, consequently, more efficient manner so that manual analysis and trial and error methods can be foregone.

The classifier, based on the neural network approach discussed in the first paragraph, currently is being prototyped for the ASTER Polar Cloud Mask algorithm deliverable. The code is being modified to conform the declaration of variables, documentation of variable names, variable naming, etc. to the C coding standards provided by the project office. The global variables are being eliminated and the code is being modularized into functions that only perform a single task. Data structures are being designed for the neural networks so that their initialization can be dynamic. This modification provides the mechanism for updating the classifier without recompilation. The code is being constructed so that PGS toolkit calls can be supplanted where appropriate (e.g., the i/o calls). The software design document is currently being prepared and should be completed early in 1997.

During this reporting period the ASTER Polar Cloud Mask Algorithm Theoretical Basis Document (ATBD) was revised and submitted to the EOS project office. The most significant change to the algorithm, since the last version of the ATBD was issued, is the replacement of the PHM with neural networks in stage 3 of the classifier. The Validation Plan and Quality Assurance Plan that were prepared earlier this year were also incorporated into the document. An anonymous ftp site was set up for GIF images of the 3-band overlay and classification mask, for the scenes presented in the aforementioned journal submission on the HNN classifier. The revised ATBD was submitted in electronic format and can be viewed at http://spso.gsfc.nasa.gov/atbd/pgl.html.

During this reporting period four reviewers provided comments on the revised version of the ATBD and Validation Plan that was submitted to the project office in August. Two major concerns were raised in the reviews. The first concern was over the time required to adapt the current version of the algorithm to the new ASTER data set. The algorithm is currently designed to derive a polar cloud mask from Landsat TM data and there are some significant differences between the ASTER and Landsat TM instruments. The second concern was over the lack of surface validation of the polar cloud mask to be derived from the ASTER data set.

To address the first concern a Post Launch Operational Plan for the ASTER Polar Cloud Mask has been developed. The purpose of the plan is to identify all the necessary efforts that must be undertaken to adapt the algorithm to the ASTER data set in a timely manner. It consists of six stages or steps which are estimated to take six months to complete from the time that the first calibrated and validated registered radiance data are available from ASTER. The six steps are: 1) development of a training set, 2) generation of the feature set, 3) feature selection, 4) classifier training, 5) classifier testing, and 6) analysis of the results.

The fundamental basis for adapting the classifier to the ASTER data set is the development of a representative set of samples from the polar regions. Approximately 200 scenes will be selected over a four-month period. From each of those scenes, 25 to 50 contiguous pixel regions (samples) will be selected and labeled. The number of labeled samples selected from each scene will be a function of the complexity of the scene (e.g., fewer samples will be selected from a completely cloud covered scene as compared to one which is heterogeneous with nearly all the classes present). These samples for all 200 scenes will total to 5000 to 10,000 samples or 250,000 to 500,000 pixel samples. Using our current software and hardware for selecting samples, about 5 scenes per day can be processed for sample extraction, which equates to 40 man days or 2 man months to process 200 scenes. The sample selection will be performed using the IVICS or SIVIS software modified to handle the ASTER 14 channel data.

As the samples are being extracted a whole host of derivative features will be computed. They include ratios, differences, arctans, and normalized differences for any pair of bands. They also include Euclidean distances and Hue/Saturation/Value for many three-way combination of bands. The feature generation software is designed so that the feature set can be dynamically selected for any combination of the 14 ASTER channels. The distributional characteristics of the features also are analyzed for correlation with ancillary information such as geographic location, season, ecosystem, etc. so that the samples are grouped accordingly. The classifier will be trained on each logical grouping of samples and, in the operational mode, the classifier will dynamically select the appropriate weights, rules, features, etc. as a function of the ancillary information.

The total number of possible features that can be generated numbers in the hundreds and using all of them in the classifier is intractable. Therefore, the total possible feature set is filtered down to no more than 40 features and, if possible, to less than 20. Several methods have been used in the past to perform this feature selection process and each has provided a set of features (although different) that result in equivalent performance. The methods include: 1) measures of divergence, 2) paired histogram overlap, 3) fuzzy logic forward and backward elimination, and 4) hypothesis testing. Several feature sets will be derived for testing to ensure that one set is not significantly better than the others.

The labeled samples will be parsed into two subsets - a training subset and a testing subset. The samples in the testing subset will be mutually exclusive of the samples in the training subset. The classifier will be trained on the training subset and tested on the testing subset. In this way a robust algorithm can be ensured. The classifier will be trained using the aforementioned feature sets. The training will be performed in both the class ambiguity reduction stage and in the leaf neural networks. Special attention will be devoted to analysis of the adaptive thresholds derived by the preclassifier.

After the classifier is trained on the labeled samples (or each set of labeled samples corresponding to a specific set of ancillary conditions), the classifier will be tested on the testing set of samples, and applied to the full scene imagery so that classification masks can be generated. The results from the testing on the samples will be assembled into confusion matrices. The classifier will be deemed adequate if the clear/cloud classification accuracy is at least 95 percent and the within clear class and within cloud class accuracy is at least 85 percent. The full scene classification masks will be inspected for performance and any major classification problems. The masks will be assembled at an FTP site for inspection by prospective users.

Although there are some significant differences between the Landsat TM and ASTER data sets, development and validation of the ASTER algorithm with Landsat TM data is a good choice. Bands 2, 3, 4, and 5 of Landsat TM are nearly identical to ASTER bands 1, 2, 3, and 4 in spectral location and spectral resolution. ASTER bands 1-3 have higher spatial resolution than Landsat TM (15 m vs. 30 m) which should provide for at least as good cloud masking performance and potentially better. ASTER band 4 has the same spatial resolution as Landsat TM band 5. Landsat TM band 7 is equivalent to the integration of ASTER bands 5-9. The spatial resolution is nearly identical to that of Landsat TM. The availability of the five higher spectral resolution ASTER bands could improve performance since some soil and rock absorption features appear in this spectral region, whereas the reflectance of clouds is relatively uniform. Landsat TM band 6 is equivalent to the integration of ASTER bands 10-14. The spatial resolution of the ASTER bands is higher (90 m vs. 120 m) and could improve performance. The availability of the five higher spectral resolution ASTER TIR bands also could improve performance, again because some soil and rock absorption features appear in this spectral region. Detection of thin cirrus could improve over Landsat TM as the absorption path below the clouds is different across this spectral region and makes detection of these transparent cloud possible by using thermal IR band differences. Potentially, the most significant difference in the signal between the two instruments are the noise, sensitivity, and calibration accuracy. As an interim classifier (during the six month post launch algorithm adaptation period), the ASTER data could be transformed into a Landsat TM equivalent signal and the algorithm could be applied in its current form. It should be possible to get at least as good performance at launch with the current algorithm applied to ASTER as is being obtained now with Landsat TM. The post launch retraining of the classifier should only improve performance.

There is no other data set that remotely approaches the variety of surface and atmospheric conditions, or spatial and temporal variability over the polar regions, with the similar spatial and spectral characteristics of ASTER, than does the Landsat TM data set (albeit the limited coverage of the Landsat TM data set). It would take 7000 Landsat TM scenes to cover the polar regions (poleward of 60 degrees) once. If we assumed good representation of the polar regions for validation, spatially and temporally, was one seventh of the spatial area, 12 times per year, and 4 different years, then 1000 * 12 * 4 = 48,000 scenes would be required. This is certainly not practical; therefore, using the most variety in terms of spatial and temporal coverage is the best alternative.

Between now and launch, as part of the ongoing prelaunch validation effort, the classifier will be tested on additional datasets. The algorithm will be tested on an additional 60 Landsat TM scenes (over and above the 82 tested to date). These additional scenes include representative regions of sea ice, ice/snow, mountains, forests, tundra, wetlands, shadows (cloud and topographic), and various cloud types (many very thin and transparent). The algorithm will also be tested on AVIRIS data. The AVIRIS bands can be averaged to simulate ASTER bands 1-9; however, AVIRIS does not have any thermal IR bands. The ability to validate the algorithm using AVIRIS is limited due to lack of thermal IR bands and polar imagery. The availability of one or more thermal IR bands is important to the performance of the classifier. TIMS data will also be used in the prelaunch validation process. Five of the six TIMS bands can be used as surrogates for the five ASTER TIR bands; however, the use of TIMS in the validation of the algorithm is limited due to lack of visible, near-IR, SW-IR channels and polar scenes not only because channels in the solar wavelengths are critical to the performance of the algorithm but also because it is difficult to accurately label samples without them. MAS data was obtained over the Beaufort Sea during the summer of 1995 and that data also can be used in algorithm prelaunch validation. Many of the MAS bands can be used as surrogates for the ASTER bands. Especially important is the availability of the thermal IR bands which are important for nighttime algorithm testing. The major limitation of this dataset is its very small coverage of the polar regions both spatially and temporally in comparison to the Landsat TM dataset (only one small geographic location over a few days).

To address the concern over surface validation of the ASTER Polar Cloud Mask, 26 sites located in the polar regions have been identified for comparative measurements. A qualitative comparison will be made between surface observations and the Polar Cloud Mask at these sites. For example, if a surface observer indicates that clouds are present in a specific geographical area and at a specific time -- Does the ASTER Polar Cloud Mask indicate the presence of clouds in that area at the same time? If the type of cloud cover is broken and rapidly changing, the time of observation from both the surface and satellite is critical. Comparisons of cloud fraction will be difficult since the surface observer and the satellite are "looking" at a different field of view. For sites which have ceilometers, a single pixel in the ASTER Polar Cloud Mask can be validated. A more comprehensive validation opportunity will become available during WINCE and, FIRE III-1 and FIRE III-2 (Apr-Jun 98 and Aug-Sep 98, respectively) in Alaska.

Mission	Date	Location	Purpose
WINCE	Feb 97	Great Lakes So. Canada	Cloud detection and properties over snow-ice covered surfaces
FIRE III-1	Apr-Jun 98	Alaska	Arctic stratus over sea ice
FIRE III-2	Aug-Sep 98	Alaska	Arctic stratus over sea ice

The following is a list of the locations for the surface validation sites:

Location	Latitude/Longitude	
Barrow, Alaska	71.20N/156.50W	
Ny Alesund, Spitsbergen	79N/12E	
Georg Von Neumayer, Antarctica	70S/08W	
Syowa Base, Antarctica	69S/39E	
Bratt's Lake, Canada	50N/104W	winter
Toravere Observatory, Estonia	58N/26E	
Boulder, CO	40.13N/105.24W	winter
Franz Josef Land, Russia	80N/55E	
Dutch Harbor, Unalaska	55N/167W	
Juneau, Alaska	57W/134W	
Anchorage, Alaska	61.10N/150.01W	
Nome, Alaska	64.30N/165.26W	
Prudhoe Bay, Alaska	70.15N/148.20W	
Aklavik Airport, Canada	68.13N/135.00W	

Location	Latitude/Longitude
Alert Airport, Canada	82.31N/62.17W
Baker Lake Airport, Canada	64.18N/96.05W
Yellowknife Airport, Canada	62.28N/114.27W
Godthak, Greenland	64.12N/51.41W
Kulusuk, Greenland	65.34N/37.07W
Reykjavik, Iceland	64.08N/21.54W
Tromso/Langues, Norway	69.41N/18.55E
Murmansk, Russia	68.58N/33.03E
Byrd Base, Antarctica	80.01S/119.32W
McMurdo Base, Antarctica	77.51S/166.40E
Palmer Base, Antarctica	64.46S/64.05W
Siple Base, Antarctica	75.55S/83.55W

Some other issues were raised by the reviewers. One reviewer noted that there is another ASTER cloud mask and was concerned about redundancy. The ASTER Polar Cloud Mask is generated only on request for polar scenes while the other cloud mask is applied to every ASTER scene. The Polar Cloud Mask is a data product for public distribution while the other mask is for internal ASTER product generation use only and is used to determine which scenes are cloud free and retained for further processing. The other cloud mask is faster but not as accurate. The Polar Cloud Mask algorithm could be trained on global data sets and used as a validation tool for MODIS and CERES products. It could be used to detect aerosols, smoke, and shadows which is especially important for validation of those products. Another reviewer expressed concern that the methodology for ASTER Polar Cloud Mask product is not the same as the methodology for CERES Global Cloud Mask product. The ASTER Polar Cloud Mask uses adaptive thresholding as a front-end for several back propagation neural networks in a hierarchical configuration. The CERES Global Cloud Mask uses paired histogram classification. Testing revealed that the paired histogram and neural networks provided equivalent performance when applied to AVHRR data sets (the CERES Cloud Mask algorithm development surrogate) while neural networks out-performed the paired histogram approach when applied to Landsat TM data sets. We concluded that the higher spatial resolution Landsat TM data provides for much higher spectral variability among the classes than does the lower spatial resolution AVHRR data. We suspect that the neural networks are better able to generalize when distributional overlap among classes is higher like seen in high spatial resolution imagery. Another reviewer wanted to know what was to be done for the nighttime algorithm. The current daytime Polar Cloud Mask algorithm has been applied to scenes with solar zenith angles as high as 81 degrees. We anticipate that scenes obtained at higher solar zenith angles will require a nighttime algorithm. It has been estimated that 90 percent of the imagery collected by ASTER will be during the daytime and only 10 percent at night, so the major developmental effort has been devoted to the daytime algorithm. However, the development of a nighttime algorithm is problematic as we really do not have any ASTER-like polar nighttime data available. Although TIMS and MAS have good surrogate bands, both data sets were obtained during the daytime and only MAS has been used to obtain any polar region data and calibrated and processed data is yet to become available. Fewer classes will be detected in the nighttime product as the contrast between classes in the TIR bands is much smaller than it is in the solar wavelength bands and cloud detection is much more difficult.

A conference paper describing the ASTER polar cloud mask was presented at the International Symposium on Optical Science, Engineering, and Instrumentation, SPIE's Annual Meeting, held 4-9 August 1996 in Denver, CO. An oral presentation was made on August 6th during the Infrared Spaceborne Remote Sensing IV Session.

Dr. Kwo-Sen Kuo traveled to Japan in December to attend the international ASTER Science Team meeting. During that meeting he presented a briefing on the status of the Polar Cloud Mask development effort and on many of the issues discussed above. Dr. Ronald M. Welch attended the EOS Validation Workshop, also in December, in which he presented a briefing addressing the aforementioned concerns of the reviewers.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	7 January 1997	Semi-annual, 1 July-31 December 1996

4. TITLE AND SUBTITLE
The Effect of Cloud Inhomogeneities Upon Radiative Fluxes, and the Supply of a Cloud Truth Validation Dataset

5. FUNDING NUMBERS
C - NAS5-31718

6. AUTHOR(S)
Ronald M. Welch

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute of Atmospheric Sciences
S.D. School of Mines and Technology
501 E. St. Joseph Street
Rapid City, SD 57701-3995

8. PERFORMING ORGANIZATION REPORT NUMBER
SA-10

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt Road
Greenbelt, MD 20771

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The ASTER polar cloud mask algorithm is currently under development. Several classification techniques have been developed and implemented. The merits and accuracy of each have been examined. The classification techniques investigated include fuzzy logic, hierarchical neural network, and a pairwise histogram comparison scheme based on sample histograms called the Paired Histogram Method. Scene adaptive methods also were investigated as a means to improve classifier performance. The feature, arctan of Band 4 and Band 5, and the Band 2 vs. Band 4 feature space are key to separating frozen water (e.g., ice/snow, slush/wet ice, etc.) from cloud over frozen water, and land from cloud over land, respectively. A total of 82 Landsat TM circumpolar scenes are being used as a basis for algorithm development and testing. Numerous spectral features are being tested and include the 7 basic Landsat TM bands, in addition to ratios, differences, arctans, and normalized differences of each combination of bands. The classifier of choice for the cloud mask is a two-stage approach. In the first stage the class ambiguity for each pixel feature vector is reduced from 10 classes to less than 10 classes through the use of some simple adaptive and fixed thresholds. The feature vector is then passed on to the second stage consisting of a set of specially trained backpropagation neural networks. Each network is trained to resolve specific ambiguities and is selected accordingly. The algorithm is currently being prototyped in C.

A Post Launch Operational Plan has also been developed which outlines the process for adapting the current algorithm (designed for Landsat TM) to the ASTER data stream. There are some significant differences between Landsat TM and ASTER and the algorithm will need to be adapted to the additional ASTER channels not present in Landsat TM. A validation plan has also been developed and includes 26 surface validation sites.

14. SUBJECT TERMS
ASTER, Landsat TM, polar cloud mask, scene classification, neural network

15. NUMBER OF PAGES
10

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	